**RT-MAE 9726**

***IMPROVED MAXIMUM LIKELIHOOD ESTIMATION IN ONE-PARAMETER EXPONENTIAL FAMILY MODELS***

*by*

*Gauss M. Cordeiro, Silvia L.P. Ferrari, Denise A.Botter*
*and*
*Francisco Cribari-Neto*

**Palavras-Chave:** Asymptotic expansion; Bartlett-type correction; Edgeworth expansion; Exponential family; Maximum likelihood estimation; Standardized maximum likelihood estimate.

**Classificação AMS: 62F11, 62F10, 62E25.**
**(AMS Classification)**

**- Outubro de 1997 -**

# IMPROVED MAXIMUM LIKELIHOOD ESTIMATION IN ONE–PARAMETER EXPONENTIAL FAMILY MODELS*

**Gauss M. Cordeiro**
*Departamento de Estatística*
*Universidade Federal de Pernambuco, Brazil*

**Silvia L. P. Ferrari**
*Departamento de Estatística*
*Universidade de São Paulo, Brazil*

**Denise A. Botter**
*Departamento de Estatística*
*Universidade de São Paulo, Brazil*

**Francisco Cribari–Neto**
*Department of Economics*
*Southern Illinois University, USA*

**Abstract**

We propose a new statistic which is a function of the maximum likelihood estimate of a scalar parameter $\theta$ and whose distribution is standard normal excluding terms of order $\mathcal{O}(n^{-3/2})$ and smaller, where $n$ is the sample size. The proposed statistic is a polynomial transformation of the classical maximum likelihood estimate of at most degree 3. We apply our main result to the one-parameter exponential family model and to a number of special distributions of this family. Some simulation results illustrate the superiority of our statistic over the usual standardized maximum likelihood estimate with regard to second-order asymptotic theory.

*Keywords*: Asymptotic expansion; Bartlett-type correction; Edgeworth expansion; Exponential family; Maximum likelihood estimation; Standardized maximum likelihood estimate.

## 1 Introduction

In general one-parameter models, the probability or density function is indexed by an unknown scalar parameter $\theta \in \Theta$ and can be written as

$$\pi(y;\theta) = \exp\{t(y,\theta)\}, \tag{1}$$

where $\Theta$ is an open subset of $\mathbb{R}$. Let $y$ be the data vector of $n$ observations, which are independent and identically distributed with total likelihood function $L(\theta) = L(\theta; y)$. We assume that $L(\theta)$ is continuously four times differentiable with respect to $\theta \in \Theta$ and that the derivative $\mathrm{d}/\mathrm{d}\theta$ and the expectation $\mathrm{E}_\theta$ with respect to $\pi(y;\theta)$ are interchangeable. We further assume that some conditions on the smoothness of $L(\theta; y)$ and its derivatives with respect to $\theta$ hold. The maximum likelihood estimate (MLE) $\hat{\theta}$ of $\theta$, assumed to be unique when the number of observations ($n$) is large, is defined as the value of $\theta$ that satisfies $\mathrm{d}l(\theta)/\mathrm{d}\theta = 0$, where $l(\theta) = l(\theta; y)$ is the log-likelihood function. A consistent solution $\hat{\theta}$ for all $y$ is required for this equation. If there are multiple solutions to $\mathrm{d}l(\theta)/\mathrm{d}\theta = 0$, it is assumed that the consistent one is known. The obvious difficulty with nonlinear MLEs is that they cannot be expressed as explicit functions of the data.

* *Correspondence to*: Gauss M. Cordeiro, Departamento de Estatística, UFPE, Cidade Universitária, Recife/PE, 50740-540, Brazil.

To circumvent this problem we usually use iterative techniques to derive approximate solutions to the exact MLEs.

Despite substantial advances in developing higher order asymptotic theory for transformations of test statistics, as for example Bartlett-type corrected statistics (see Cribari–Neto and Cordeiro, 1996), only few attempts have been made to develop higher order asymptotic theory for transformations of the MLE in general statistical models. Our main purpose here is to define a new transformed statistic $S^* = S^*(\hat{\theta}, \theta)$ as a function of the MLE $\hat{\theta}$ and $\theta$ whose distribution is standard normal excluding terms of order $O(n^{-3/2})$. In this paper, following similar arguments of Cordeiro and Ferrari (1991), who gave a general formula of Bartlett-type corrections for test statistics whose asymptotic expansion is a finite linear combination of chi-squared distributions with suitably defined number of degrees of freedom, we obtain a new kind of Bartlett-type correction for a class of test statistics which converge in distribution to a standard normal distribution.

Some similar results of higher-order refinements of the MLEs have already been considered by Barndorff-Nielsen (1983, 1988, 1990), Jensen (1992) and Ferrari et al (1996). It is useful to briefly review these results. Barndorff-Nielsen (1988) derived the $p^*$ formula, which provides an approximation to the conditional density of $\hat{\theta}$, given an ancillary statistic $a$. The conjunction of $a$ and $\hat{\theta}$ constitutes a sufficient statistic and $a$ is distribution constant in the sense adopted in Barndorff-Nielsen (1983), either exactly or to the requisite order of approximation. The $p^*$-formula is defined by

$$p^*(\hat{\theta}; \theta \mid a) = c \left\{ \frac{L(\theta; \hat{\theta}, a)}{L(\hat{\theta}; \hat{\theta}, a)} \right\} |j(\hat{\theta}, a)|^{1/2}. \qquad (2)$$

It is important to note that the likelihood function $L(\theta; \hat{\theta}, a)$ in (2) has been expressed as a function of $(\hat{\theta}, a)$ in its dependence on the data, $|j(\hat{\theta}, a)|$ is the determinant of the observed Fisher information for $\theta$ evaluated at $\theta = \hat{\theta}$ and which is a function of $\hat{\theta}$ and $a$, and $c = c(\theta, a)$ is a normalizing constant chosen so that the total integral of (2) with respect to $\hat{\theta}$, holding $a$ constant, equals one. Equation (2) also holds when $\theta$ is a parameter vector.

When $\theta$ is scalar, there are also available accurate approximations to the distribution function of $\hat{\theta}$ which may be more useful in applications. In fact, Barndorff-Nielsen (1988, 1990) derived from (2) the general tail area approximation

$$F(\hat{\theta} \mid a; \theta) = \left\{ \Phi(r) + \phi(r) \left( \frac{1}{r} - \frac{1}{u} \right) \right\} \{1 + O(n^{-3/2})\}, \qquad (3)$$

where

$$r = \pm \left[ 2 \left\{ \log \left( \frac{L(\hat{\theta})}{L(\theta)} \right) \right\} \right]^{1/2}$$

is the signed log likelihood ratio statistic and

$$u = \left\{ \frac{\partial l(\theta; \hat{\theta}, a)}{\partial \theta} \Big|_{\theta=\hat{\theta}} - \frac{\partial l(\theta; \hat{\theta}, a)}{\partial \theta} \right\} |j(\hat{\theta}, a)|^{-1/2}$$

can be viewed as a test statistic which is analogous – and in a sense dual – to the score statistic. Here, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions, respectively. An alternative version of (3) was proposed by Barndorff-Nielsen (1990) and Jensen (1992) which can be written as

$$F(\hat{\theta} \mid a; \theta) = \Phi(r^*)\{1 + O(n^{-3/2})\} ,$$

where $r^* = r + r^{-1}\log(u/r)$. The main difficulty in using equations (2) and (3) in practice lies in obtaining the ancillary statistic $a$. Our approach for computing improved significance levels or confidence intervals for $\theta$ in a wide variety of uniparametric models is more useful than the result in (3) since it does not require knowledge of any ancillary statistic.

More recently, Ferrari et al. (1996) obtained second- and third-order bias-corrected MLEs in general uniparametric models (1) and compared the corrected estimates and the usual MLE $\hat\theta$ on the basis of mean squared errors.

In this paper, we offer a simple way of correcting the distribution of the MLE to an error of order $O(n^{-3/2})$. The derivatives of the log-likelihood function are denoted by $U_\theta = \mathrm{d}l(\theta)/\mathrm{d}\theta$, $U_{\theta\theta} = \mathrm{d}^2 l(\theta)/\mathrm{d}\theta^2$, etc., and we use the following notation for their moments (Lawley, 1956): $\kappa_{\theta\theta} = \mathrm{E}(U_{\theta\theta})$, $\kappa_{\theta\theta\theta} = \mathrm{E}(U_{\theta\theta\theta})$, $\kappa_{\theta,\theta} = \mathrm{E}(U_\theta^2)$, $\kappa_{\theta,\theta\theta} = \mathrm{E}(U_\theta U_{\theta\theta})$, $\kappa_{\theta\theta,\theta\theta} = \mathrm{E}(U_{\theta\theta}^2) - \kappa_{\theta\theta}^2$, etc. We denote the derivatives of the moments as $\kappa_{\theta\theta}^{(\theta)} = \mathrm{d}\kappa_{\theta\theta}/\mathrm{d}\theta$, etc. All $\kappa$'s refer to a total over the components of $y$ and are, in general, of order $O(n)$. Under regularity conditions on $L(\theta; y)$ (Cox and Hinkley, 1974, Section 9.1), it follows that the score function $U_\theta$ is asymptotically distributed as $\mathrm{N}(0, \kappa_{\theta,\theta})$, so that the asymptotic distribution of $\hat\theta$ is $\mathrm{N}(\theta, \kappa_{\theta,\theta}^{-1})$, with an error of order $O(n^{-1/2})$.

We shall work with the standardized statistic $S = (\hat\theta - \theta)\kappa_{\theta,\theta}^{1/2}$ as a pivot function for $\theta$ which is frequently used to test the null hypothesis $H_0 : \theta = \theta_0$, or to construct confidence limits for $\theta$. The normal approximation for $S$ is usually unsatisfactory in small samples because the exact and asymptotic distributions of $S$ differ by a term of order $O(n^{-1/2})$. It is then desirable to improve on this result by adjusting $S$ so that the normal approximation holds with smaller error. To that end, we shall obtain in Section 2 a transformed statistic $S^*$ which is distributed as $\mathrm{N}(0,1)$ to an error of order $O(n^{-3/2})$ by using the Edgeworth expansion for the distribution function of $S$. The statistic $S^*$ takes the form of a polynomial of third degree in the MLE $\hat\theta$, i.e., $S^* = S + \sum_{i=0}^{3} \alpha_i S^i$, where the $\alpha$'s depend on the model under consideration solely in terms of moments $\kappa$'s of some log-likelihood derivatives up to the fourth order. The proposed statistic $S^*$ is quite general and can be used to improve the MLE for several distributions to an order of accuracy of $O(n^{-1})$. In Section 3 we obtain the modified MLE for one-parameter exponential family models. In Section 4 we present a number of special cases thus showing that our main result covers a wide range of important distributions. Section 5 considers the modified statistic $S^*$ for some classes of variance functions in natural exponential family models. Finally, Section 6 gives some simulation results which favor $S^*$ over $S$.

## 2 An adjusted pivotal quantity

It will now be seen how the statistic $S^*$ arises from the Edgeworth asymptotic expansion for the distribution function of $S$ to order $O(n^{-1})$. The Edgeworth asymptotic expansion for the distribution function of $S = (\hat\theta - \theta)\kappa_{\theta,\theta}^{1/2}$ is given, to order $O(n^{-1})$ by (see, e.g., Hill and Davis, 1968)

$$F_S(x) = \Phi(x) - \phi(x)\left\{\frac{6\eta_1 + \eta_3 h_2(x)}{6} + \frac{360\eta_2 h_1(x) + 30\eta_4 h_3(x) + \eta_6 h_5(x)}{24}\right\}, \tag{4}$$

where $\eta_1 = \kappa_{\theta,\theta}^{1/2} b_1(\theta) + O(n^{-3/2})$, $\eta_2 = \kappa_{\theta,\theta}\{v_2(\theta) + b_1(\theta)^2\} + O(n^{-2})$, $\eta_3 = \rho_{3\hat\theta} + O(n^{-3/2})$, $\eta_4 = \rho_{4\hat\theta} + 4\rho_{3\hat\theta}\kappa_{\theta,\theta}^{1/2} b_1(\theta) + O(n^{-2})$ and $\eta_6 = 10\rho_{3\hat\theta}^2 + O(n^{-2})$ with $b_1(\theta)$ and $v_2(\theta)$ being the $n^{-1}$ and $n^{-2}$ terms in the bias and variance of $\hat\theta$, respectively, and $\rho_{3\hat\theta}$ and $\rho_{4\hat\theta}$ being the third and fourth cumulants of $\hat\theta$ which are of orders $n^{-1/2}$ and $n^{-1}$, respectively. The polynomials appearing in (4) are the standard Hermite polynomials given by $h_1(x) = x$, $h_2(x) = x^2 - 1$, $h_3(x) = x^3 - 3x$ and $h_5(x) = x^5 - 10x^3 + 15x$. The $\eta$'s are functions of the moments $\kappa$'s. Formulae for $b_1(\theta)$, $v_2(\theta)$, $\rho_{3\hat{\theta}}$ and $\rho_{4\hat{\theta}}$ are given by Shenton and Bowman (1977, Sections 2.7.6 and 2.7.7). The two correction terms in (4) are of orders $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(n^{-1})$, respectively. Equation (4) holds provided that suitable regularity conditions are verified (Feller, 1971; Bhattacharya and Rao, 1976; Bhattacharya and Ghosh, 1978; Skovgaard, 1981a,b).

The form (4) of the distribution function of $S$ containing terms of order $n^{-1/2}$ and $n^{-1}$ suggests the use of a modified statistic defined as

$$S^* = S - c_1(S) - c_2(S), \tag{5}$$

where $c_1(S)$ and $c_2(S)$ are additive stochastic correction quantities of orders $\mathcal{O}_p(n^{-1/2})$ and $\mathcal{O}_p(n^{-1})$, respectively, which are functions of the pivot $S$. The functions $c_1(S)$ and $c_2(S)$ are now determined to make the distribution of $S^*$ to order $n^{-1}$, say $F_{S^*}(x)$, identical to $\Phi(x)$. Given that the two terms in braces in (4) are terms of orders $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(n^{-1})$ and that $c_1(S)$ and $c_2(S)$ are assumed to be of orders $\mathcal{O}_p(n^{-1/2})$ and $\mathcal{O}_p(n^{-1})$, respectively, the distribution of $S^*$ can be derived from formulae (4) and (5) by using the approach developed by Cordeiro and Ferrari (1997). We obtain

$$F_{S^*}(x) = F_S(x) + c_1(x)f_S(x) + c_2(x)f_S(x) + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\{c_1(x)^2 f_S(x)\} + \mathcal{O}(n^{-3/2}).$$

The functions $c_1(x)$ and $c_2(x)$ are then formed from (4) and the above equation by collecting terms that are of equal order in $n^{-1/2}$ and $n^{-1}$. We obtain a polynomial of third degree in the pivot $S$

$$S^* = S + \sum_{i=0}^{3} \alpha_i S^i, \tag{6}$$

where $\alpha_0 = (\eta_3 - 6\eta_1)/6$, $\alpha_1 = (36\eta_1^2 - 36\eta_2 - 12\eta_1\eta_3 - 14\eta_3^2 + 9\eta_4)/72$, $\alpha_2 = -\eta_3/6$ and $\alpha_3 = (12\eta_1\eta_3 + 8\eta_3^2 - 3\eta_4)/72$. Note that $\alpha_0$ and $\alpha_2$ are $\mathcal{O}(n^{-1/2})$ while $\alpha_1$ and $\alpha_3$ are $\mathcal{O}(n^{-1})$.

Using the general formulae for $b_1(\theta)$, $v_2(\theta)$, $\rho_{3\hat{\theta}}$ and $\rho_{4\hat{\theta}}$ given by Shenton and Bowman (1977, equations (2.30a,b), (2.31a,b)) and some Bartlett identities, which usually simplify the computation of the $\kappa$'s, Cordeiro and Ferrari (1997) obtained the coefficients $\alpha$'s as functions of the moments of the joint distribution of certain derivatives of the log-likelihood function:

$$\begin{aligned}
\alpha_0 &= (2\kappa_{\theta\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)})/(6\kappa_{\theta,\theta}^{3/2}),\\
\alpha_1 &= (\kappa_{\theta\theta\theta\theta} - \kappa_{\theta\theta,\theta\theta} - 2\kappa_{\theta\theta}^{(\theta\theta)})/(8\kappa_{\theta,\theta}^2) - (9\kappa_{\theta\theta}^{(\theta)2} + 6\kappa_{\theta\theta}^{(\theta)}\kappa_{\theta\theta\theta} - 8\kappa_{\theta\theta\theta}^2)/(36\kappa_{\theta,\theta}^3),\\
\alpha_2 &= (\kappa_{\theta\theta\theta} - 3\kappa_{\theta\theta}^{(\theta)})/(6\kappa_{\theta,\theta}^{3/2}),\\
\alpha_3 &= -(\kappa_{\theta\theta\theta\theta} - 4\kappa_{\theta\theta\theta}^{(\theta)} + 6\kappa_{\theta\theta}^{(\theta\theta)} + 3\kappa_{\theta\theta,\theta\theta})/(24\kappa_{\theta,\theta}^2) - (3\kappa_{\theta\theta}^{(\theta)} - 2\kappa_{\theta\theta\theta})\kappa_{\theta\theta\theta}/(18\kappa_{\theta,\theta}^3).
\end{aligned} \tag{7}$$

After computing the $\alpha$'s from equations (7) for the model under investigation, the improved statistic $S^*$ for the pivot $S = (\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$ follows immediately from (6). Then, in wide generality, the new pivotal quantity $S^*$ is asymptotically (standard) normally distributed with error which is typically $\mathcal{O}(n^{-3/2})$. The finite-sample behavior of $S^*$ and $S$ can be quite different. In fact, simulation results reported in Section 6 favor $S^*$ over $S$.

In view of the discussion above, it appears that tests based on $S^*$ can be viewed as meaningful improvements of tests based on $S$. Formulae for $S^*$ provides the basis for obtaining the corrected version of the MLE $\hat{\theta}$. It is easy to check that $\hat{\theta}^* = \hat{\theta} + \sum_{i=0}^{3} \alpha_i(\hat{\theta}-\theta)^i \kappa_{\theta,\theta}^{(i-1)/2}$ follows a $\mathrm{N}(\theta, \kappa_{\theta,\theta}^{-1})$ distribution with the error of the approximation typically being $\mathcal{O}(n^{-2})$. The above polynomial transformation for $\hat{\theta}^*$ looks very much like a truncated power series in the pivot $\hat{\theta}-\theta$. It follows that $S^* = (\hat{\theta}^* - \theta)\kappa_{\theta,\theta}^{1/2} \sim \mathrm{N}(0,1) + \mathcal{O}_p(n^{-3/2})$ implies that $\{\theta : |\hat{\theta}^* - \theta|\kappa_{\theta,\theta}^{1/2} \leq z\}$ is an improved set of approximate confidence intervals for $\theta$, where $z$ is a normal upper point, i.e., values of $\theta$ outside this set are incompatible with the data.

It should be noted that the theoretical results presented in this section are only valid for continuous distributions. In view of this, the behavior of the statistic $S^*$ for discrete models should be examined by simulation studies.

# 3 One-parameter exponential family models

In this section we obtain simple expressions for the coefficients $\alpha$'s in (6) that can be used to improve the standardized MLE $S = (\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$ in one-parameter exponential family models. We apply the polynomial transformation (6) to the statistic $S$ to obtain a corrected statistic $S^*$ such that $\mathrm{P}(S^* \leq x) = \Phi(x) + \mathcal{O}(n^{-3/2})$ whereas $\mathrm{P}(S \leq x) = \Phi(x) + \mathcal{O}(n^{-1/2})$. We have chosen this family of models because it enjoys wide application and many useful mathematical properties. Moreover, the formulae derived here for the $\alpha$'s are algebraically more appealing for applications than equations (7) and can be readily used by applied researchers since they only require trivial operations on suitably defined functions and their derivatives. In Section 4, we present a number of special cases thus showing that our formulae have a wide range of important applications. Recently, Cordeiro et al. (1995) and Ferrari et al. (1996) derived Bartlett-type corrections to the likelihood ratio and score test statistics, respectively, for the class of one-parameter exponential family models.

The one-parameter exponential family model admits the probability or density function

$$\pi(y;\theta) = \frac{1}{\zeta(\theta)} \exp\{-\alpha(\theta)d(y) + v(y)\}, \tag{8}$$

where $\theta$ is a scalar parameter of interest, $\zeta(\cdot)$, $\alpha(\cdot)$, $d(\cdot)$ and $v(\cdot)$ are known functions, and $\zeta(\cdot)$ is positive-valued. We also assume that the support set of the family (8) is independent of $\theta$ and that $\alpha(\cdot)$ and $\zeta(\cdot)$ have continuous first four derivatives, $\mathrm{d}\alpha(\theta)/\mathrm{d}\theta$ and $\mathrm{d}\beta(\theta)/\mathrm{d}\theta$ being different from zero for all $\theta$ in the parameter space, where $\beta(\theta) = -\mathrm{E}\{d(y)\} = \zeta'(\theta)/\{\zeta(\theta)\alpha'(\theta)\}$, with primes denoting derivatives with respect to $\theta$. It then follows that $\mathrm{var}\{d(y)\} = \beta'/\alpha'$, $\kappa_{\theta,\theta} = -\kappa_{\theta\theta} = n\alpha'\beta'$, $\kappa_{\theta\theta\theta} = -n(2\alpha''\beta' + \alpha'\beta'')$, $\kappa_{\theta\theta\theta\theta} = -3n(\alpha'''\beta' + \alpha''\beta'') - n\alpha'\beta'''$, etc. Using these results, it is possible to simplify the expressions for the $\alpha$'s in equations (7). After some algebra, we obtain

$$\begin{aligned}
\alpha_0 &= (\alpha'\beta'' - \alpha''\beta')/(6\alpha'\beta'\sqrt{n\alpha'\beta'}), \\
\alpha_1 &= \{-9\alpha'''\alpha'\beta'^2 + 13\alpha''^2\beta'^2 + \alpha''\alpha'\beta''\beta' - \alpha'^2(14\beta''^2 - 9\beta'''\beta')\}/(72n\alpha'^3\beta'^3), \\
\alpha_2 &= (\alpha''\beta' + 2\alpha'\beta'')/(6\alpha'\beta'\sqrt{n\alpha'\beta'}), \\
\alpha_3 &= \{3\alpha'''\alpha'\beta'^2 - \alpha''^2\beta'^2 + 5\alpha''\alpha'\beta''\beta' - \alpha'^2(4\beta''^2 - 9\beta'''\beta')\}/(72n\alpha'^3\beta'^3).
\end{aligned} \tag{9}$$

Some features of equations (9) are noteworthy. First, the $\alpha$'s depend on the model only through the functions $\alpha$ and $\beta$ and their first three derivatives with respect to $\theta$. Second, the improved statistic $S^*$ in (6) is now very easy to compute for any exponential family model. Third, when $\alpha$ equals $\theta$, which corresponds to the natural exponential family, the $\alpha$'s reduce to $\alpha_0 = \gamma_1/(6\sqrt{n})$, $\alpha_1 = (9\gamma_2 - 14\gamma_1^2)/(72n)$, $\alpha_2 = 2\alpha_0$ and $\alpha_3 = \alpha_1$, where $\gamma_1^2 = \beta''^2/\beta'^3$ and $\gamma_2 = \beta'''/\beta'^2$ are the third and fourth standardized cumulants of $-d(y)$, respectively. Fourth, by entering equations (9) into a computer algebra system such as MATHEMATICA (Wolfram, 1996) or MAPLE (Abell and Baselton, 1994), one can obtain the $\alpha$'s as functions of $\theta$ with minimal effort (see Section 4). Although the calculation of the $\alpha$'s is straightforward for any model it is rather difficult to explain their general structure in (9). The main difficulty in interpreting equations (9) is that the individual terms are not invariant under reparameterization and therefore they have no geometric interpretation which is independent of the coordinate system chosen.

# 4 Special cases

In this section, we use equations (9) to derive improved statistics of the type (6) for a number of important distributions that belong to the one-parameter exponential family. We consider 24 special distributions and give closed-form expressions for the $\alpha$'s. The calculations were done using MATHEMATICA and MAPLE. Distributions (i) through (viii) involve discrete random variables whereas continuous random variables are considered in cases (ix) through (xxiv). The special distributions listed below have a wide range of practical applications in various fields such as engineering, biology, medicine, economics, among others (Johnson and Kotz and Balakrishnan, 1994, 1995; Johnson, Kotz and Kemp, 1992). The distributions are:

(i) Binomial ($0 < \theta < 1$, $m \in \mathbb{N}$, $m$ known, $y = 0, 1, 2, \ldots, m$): $\alpha(\theta) = -\log\{\theta/(1-\theta)\}$, $\zeta(\theta) = (1-\theta)^{-m}$, $d(y) = y$, $v(y) = \log\binom{m}{y}$, $\hat{\theta} = \bar{y}/m$,

$$\alpha_0 = \frac{-1 + 6\theta - 14\theta^2 + 16\theta^3 - 9\theta^4 + 2\theta^5}{6\{m\,n\,\theta(1-\theta)^9\}^{1/2}}, \quad \alpha_1 = \frac{-5 + 2\theta - 2\theta^2}{72m\,n\,\theta - 72m\,n\,\theta^2}, \quad \alpha_2 = -\alpha_0,$$

$$\alpha_3 = \frac{5 - 14\theta + 14\theta^2}{72m\,n\,\theta - 72m\,n\,\theta^2}.$$

(ii) Negative binomial ($0 < \theta < 1$, $\gamma > 0$, $\gamma$ known, $y = 0, 1, 2, \ldots$): $\alpha(\theta) = -\log\theta$, $\zeta(\theta) = (1-\theta)^{-\gamma}$, $d(y) = y$, $v(y) = \log\binom{\gamma+y-1}{y}$, $\hat{\theta} = \bar{y}/(\bar{y}+\gamma)$,

$$\alpha_0 = \frac{1+\theta}{6(n\,\gamma\,\theta)^{1/2}}, \quad \alpha_1 = \frac{-5 + 8\theta - 5\theta^2}{72n\,\gamma\,\theta}, \quad \alpha_2 = \frac{5\theta - 1}{6(n\,\gamma\,\theta)^{1/2}}, \quad \alpha_3 = \frac{5 - 20\theta + 53\theta^2}{72n\,\gamma\,\theta}.$$

(iii) Poisson ($\theta > 0$, $y = 0, 1, 2, \ldots$): $\alpha(\theta) = -\log(\theta)$, $\zeta(\theta) = \exp\{\theta\}$, $d(y) = y$, $v(y) = -\log(y!)$, $\hat{\theta} = \bar{y}$, $\alpha_0 = 1/\{6(n\,\theta)^{1/2}\}$, $\alpha_1 = -5/(72n\,\theta)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -\alpha_1$.

(iv) Truncated Poisson ($\theta > 0$, $y = 1, 2, \ldots$): $\alpha(\theta) = -\log(\theta)$, $\zeta(\theta) = e^{\theta}(1-e^{-\theta})$, $d(y) = y$, $v(y) = -\log(y!)$, $\hat{\theta}$ is obtained as solution of the equation $1/\{\hat{\theta}(1 - 1/e^{\hat{\theta}})\} = 1/\bar{y}$,

$$\alpha_0 = \{-1 - 3\theta - \theta^2 + e^{\theta}(2 + 3\theta - \theta^2) - e^{2\theta}\}/[6\{n\,\theta\,e^{\theta}\,(e^{\theta} - \theta - 1)^3\}^{1/2}],$$

$$\begin{aligned}\alpha_1 \;=\; & \{-5 - 12\theta - 37\theta^2 - 21\theta^3 - 5\theta^4 + e^{\theta}(20 + 36\theta + 100\theta^2 + 27\theta^3 + 8\theta^4) \\ & + e^{2\theta}(-30 - 36\theta - 89\theta^2 + 3\theta^3 - 5\theta^4) + e^{3\theta}(20 + 12\theta + 26\theta^2 - 9\theta^3) \\ & -5e^{4\theta}\}/\{72n\,\theta e^{\theta}(e^{\theta} - \theta - 1)^3\},\end{aligned}$$

$$\alpha_2 = \{1-3\theta-2\theta^2+e^\theta(-2+3\theta-2\theta^2)+e^{2\theta}\}/[6\{n\,\theta\,e^\theta\,(e^\theta-\theta-1)^3\}^{1/2}],$$

$$\begin{aligned}\alpha_3 &= \{5+\theta^2+15\theta^3+5\theta^4+e^\theta(-20+20\theta^2+27\theta^3+28\theta^4)\\ &\quad +e^{2\theta}(30-43\theta^2-33\theta^3+5\theta^4)+e^{3\theta}(-20+22\theta^2-9\theta^3)\\ &\quad +5e^{4\theta}\}/\{72n\,\theta\,e^\theta(e^\theta-\theta-1)^3\},\end{aligned}$$

(v) Logarithmic series ($0<\theta<1$, $y=1,2,\ldots$): $\alpha(\theta)=-\log(\theta)$, $\zeta(\theta)=-\log(1-\theta)$, $d(y)=y$, $v(y)=-\log(y)$, $\hat\theta$ is obtained as solution of the equation $\bar y=-\hat\theta/\{(1-\hat\theta)\log(1-\hat\theta)\}$,

$$\alpha_0=[2\theta^2+3\theta\log(1-\theta)+\{\log(1-\theta)\}^2(1+\theta)]/[-n\,\theta\,\{\theta+\log(1-\theta)\}^3]^{1/2},$$

$$\begin{aligned}\alpha_1 &= [2\theta^4+6\theta^3\log(1-\theta)+\theta^2\{\log(1-\theta)\}^2(11+20\theta)+\theta\{\log(1-\theta)\}^3(12+12\theta-9\theta^2)\\ &\quad +\{\log(1-\theta)\}^4(5-8\theta+5\theta^2)]/[72n\,\theta\{\theta+\log(1-\theta)\}^3],\end{aligned}$$

$$\alpha_2=[4\theta^2+3\theta\log(1-\theta)(1+\theta)-\{\log(1-\theta)\}^2(1-5\theta)]/[-n\,\theta\{\theta+\log(1-\theta)\}^3]^{1/2},$$

$$\begin{aligned}\alpha_3 &= -[38\theta^4+\theta^3\log(1-\theta)(66+48\theta)+\theta^2\{\log(1-\theta)\}^2(23+92\theta+24\theta^2)\\ &\quad +\theta^2\{\log(1-\theta)\}^3(24+81\theta)+\{\log(1-\theta)\}^4(5-20\theta+53\theta^2)]/[72n\,\theta\{\theta\\ &\quad +\log(1-\theta)\}^3],\end{aligned}$$

(vi) Power Series ($\theta>0$, $a_y\geq 0$, $y=0,1,2,\ldots$): $\alpha=-\log(\theta)$, $\zeta(\theta)=\sum_{y=0}^{\infty}a_y\theta^y$, $d(y)=y$, $v(y)=\log(a_y)$, $\hat\theta$ is obtained as solution of the equation $\bar y=\hat\theta g(\hat\theta)$,

$$\begin{aligned}\alpha_0 &= \frac{-(g+3\theta g'+\theta^2 g'')}{6\{n\,\theta\,(g+\theta g')^3\}^{1/2}},\\ \alpha_1 &= \{-5g^2-12\theta gg'+\theta^2(-63g'^2+26gg'')+\theta^3(-30g'g''+9gg''')\\ &\quad +\theta^4(-14g''^2+9g'g''')\}/\{72n\,\theta(g+\theta\,g')^3\},\\ \alpha_2 &= \frac{g-3\theta g'-2\theta^2 g'')}{6\{n\,\theta(g+\theta g')^3\}^{1/2}},\\ \alpha_3 &= \{5g^2+\theta^2(-21g'^2+22gg'')+\theta^3(6g'g''+9gg''')\\ &\quad +\theta^4(-4g''^2+9g'g''')\}/\{72n\,\theta(g+\theta g')^3\},\end{aligned}$$

where $g=g(\theta)=\mathrm{d}\log\zeta(\theta)/\mathrm{d}\theta$. Note that cases (iii), (iv), (v) and (vii) can be obtained from this case by simple specification of the function $g(.)$.

(vii) Zeta ($\theta>0$, $y=1,2,3,\ldots$): $\alpha(\theta)=\theta+1$, $\zeta(\theta)=\mathrm{Zeta}(\theta+1)$, $d(y)=\log(y)$, $v(y)=0$, $\hat\theta$ is obtained as solution of the equation $g(\hat\theta)=-n^{-1}\sum_{i=1}^{n}\log(y_i)$,

$$\alpha_0=\frac{g''}{6(ng'^3)^{1/2}},\ \alpha_1=\frac{9g'g'''-14g''^2}{72n\,g'^3},\ \alpha_2=2\alpha_0,\ \alpha_3=\frac{-4g''^2+9g'g'''}{72n\,g'^3},$$

where $\zeta$ is the Riemann zeta-function, *i.e.*, $\zeta(\theta)=\mathrm{Zeta}(\theta+1)=\sum_{i=1}^{\infty}i^{-(\theta+1)}$ (see, *e.g.*, Patterson, 1988) and $g=g(\theta)=\mathrm{d}\log\mathrm{Zeta}(\theta+1)/\mathrm{d}\theta$.

(viii) Non-central hypergeometric ($\theta > 0$, $m_1$, $m_2$, $r$ are known positive integers, $k_1 = \max\{0, r - m_2\} \le y \le \min\{m_1, r\} = k_2$): $\alpha(\theta) = \theta$, $\zeta(\theta) = D_0(\theta)$, $d(y) = -y$, $v(y) = \log\{\binom{m_1}{y}\binom{m_2}{r-y}\}$, $\hat\theta$ is obtained as solution of the equation $\bar y = D_1(\hat\theta)/D_0(\hat\theta)$,

$$\begin{aligned}
\alpha_0 &= \frac{2D_1^3 - 3D_0D_1D_2 + D_0^2D_3}{\sqrt{n}(-D_1^2 + D_0D_2)^{3/2}},\\
\alpha_1 &= (2D_1^6 - 6D_0D_1^4D_2 - 9D_0^2D_1^2D_2^2 + 27D_0^3D_2^3 + 20D_0^2D_1^3D_3\\
&\quad -48D_0^3D_1D_2D_3 + 14D_0^4D_3^2 + 9D_0^3D_1^2D_4 - 9D_0^4D_2D_4)/\{72n(D_1^2 - D_0D_2)^3\},\\
\alpha_2 &= \frac{2D_1^3 - 3D_0D_1D_2 + D_0^2D_3}{\sqrt{n}(-D_1^2 + D_0D_2)^{3/2}},\\
\alpha_3 &= -(38D_1^6 - 114D_0D_1^4D_2 + 99D_0^2D_1^2D_2^2 - 27D_0^3D_2^3 + 20D_0^2D_1^3D_3\\
&\quad -12D_0^3D_1D_2D_3 - 4D_0^4D_3^2 - 9D_0^3D_1^2D_4 + 9D_0^4D_2D_4)/\{72n(D_1^2 - D_0D_2)^3\},
\end{aligned}$$

where $D_p = D_p(\theta) = \sum_{y=k_1}^{k_2} y^p \binom{m_1}{y}\binom{m_2}{r-y}\exp\{\theta y\}$, $p = 0, 1, 2, 3, 4$.

(ix) Maxwell ($\theta > 0$, $y > 0$): $\alpha(\theta) = (2\theta^2)^{-1}$, $\zeta(\theta) = \theta^3$, $d(y) = y^2$, $v(y) = \log(y^2\sqrt{2/\pi})$, $\hat\theta = \{(\sum_{i=1}^n y_i^2)/(3n)\}^{1/2}$, $\alpha_0 = 2/\{3(6n)^{1/2}\}$, $\alpha_1 = -1/(54n)$, $\alpha_2 = -1/\{6(6n)^{1/2}\}$, $\alpha_3 = -\alpha_1$.

(x) Gamma ($k > 0, \theta > 0$, $y > 0$):

(a) $k$ known: $\alpha(\theta) = 1/\theta$, $\zeta(\theta) = \theta^k$, $d(y) = ky$, $v(y) = (k-1)\log(y) - \log\{\Gamma(k)\}$, $\hat\theta = \bar y$, $\alpha_0 = 1/\{3(k\,n)^{1/2}\}$, $\alpha_1 = -1/(36k\,n)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -7\alpha_1$.

(b) $\theta$ known: $\alpha(k) = 1 - k$, $\zeta(k) = \theta^{-k}\Gamma(k)$, $d(y) = \log(y)$, $v(y) = -\theta y$, $\hat k$ is obtained as solution of the equation $\psi(\hat k) = n^{-1}\log(\theta^n/\prod_{i=1}^n y_i)$,

$$\alpha_0 = \frac{\psi''(k)}{6\{n\psi'(k)^3\}^{1/2}}, \quad \alpha_1 = \frac{-14\psi''(k)^2 + 9\psi'(k)\psi'''(k)}{72n\,\psi'(k)^3}, \quad \alpha_2 = 2\alpha_0,$$

$$\alpha_3 = \frac{-4\psi''(k)^2 + 9\psi'(k)\psi'''(k)}{72n\,\psi'(k)^3},$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ are the gamma and digamma functions, respectively.

(xi) Burr system of distributions ($\theta > 0$, $b > 0$, $b$ known, $y > 0$): $\alpha(\theta) = \theta$, $\zeta(\theta) = g(\theta)/\theta$, $d(y) = -\log G(y)$, $v(y) = \log\{|\mathrm{d}\log G(y)/\mathrm{d}y|\}$, $\hat\theta = 1/\bar d$ for Burr II-VI and for Burr X-XII, $\hat\theta = 1/(\bar d + \log 2)$ for Burr VII, $\hat\theta = 1/\{\bar d + \log(\pi/2)\}$ for Burr VIII, $\alpha_0 = -1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = 2\alpha_0$, $\alpha_3 = -19\alpha_1$, where the functions $g(.)$ and $G(.)$ must be positive and $\bar d = n^{-1}\sum_{i=1}^n d(y_i)$. Different choices for $g(\theta)$ and $G(y)$ lead to different distributions; see Burr (1942). Burr I and Burr IX distributions do not belong to the exponential family.

(xii) Rayleigh ($\theta > 0$, $y > 0$): $\alpha(\theta) = \theta^{-2}$, $\zeta(\theta) = \theta^2$, $d(y) = y^2$, $v(y) = \log(2y)$, $\hat\theta = (n^{-1}\sum_{i=1}^n y_i^2)^{1/2}$, $\alpha_0 = 1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = -\alpha_0/4$, $\alpha_3 = -\alpha_1$.

(xiii) Pareto ($\theta > 0$, $k > 0$, $k$ known, $y > k$): $\alpha(\theta) = \theta + 1$, $\zeta(\theta) = (\theta k^\theta)^{-1}$, $d(y) = \log(y)$, $v(y) = 0$, $\hat\theta = [\log\{(\prod_{i=1}^n y_i)^{1/n}/k\}]^{-1}$, $\alpha_0 = -1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = 2\alpha_0$, $\alpha_3 = -19\alpha_1$.

(xiv) Weibull ($\theta > 0$, $\phi > 0$, $\phi$ known, $y > 0$): $\alpha(\theta) = \theta^{-\phi}$, $\zeta(\theta) = \theta^\phi$, $d(y) = y^\phi$, $v(y) = \log(\phi) + (\phi - 1)\log(y)$, $\hat\theta = (n^{-1}\sum_{i=1}^n y_i^\phi)^{1/\phi}$, $\alpha_0 = 1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = (\phi - 3)/(6\phi n^{1/2})$, $\alpha_3 = (12 - 6\phi + \phi^2)/(36\phi^2 n)$.

(xv) Power ($\theta > 0$, $\phi > 0$, $\phi$ known, $y > \phi$): $\alpha(\theta) = 1 - \theta$, $\zeta(\theta) = \theta^{-1}\phi^{\theta}$, $d(y) = \log(y)$, $v(y) = 0$, $\hat{\theta} = [\log\{\phi/(\prod_{i=1}^{n} y_i)^{1/n}\}]^{-1}$, $\alpha_0 = -1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = 2\alpha_0$, $\alpha_3 = -19\alpha_1$.

(xvi) Laplace ($\theta > 0$, $-\infty < k < \infty$, $k$ known, $y > 0$): $\alpha(\theta) = \theta^{-1}$, $\zeta(\theta) = 2\theta$, $d(y) = |y - k|$, $v(y) = 0$, $\hat{\theta} = n^{-1}\sum_{i=1}^{n} |y_i - k|$, $\alpha_0 = 1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -7\alpha_1$.

(xvii) Extreme value ($-\infty < \theta < \infty$, $\phi > 0$, $\phi$ known, $-\infty < y < \infty$): $\alpha(\theta) = \exp\{\theta/\phi\}$, $\zeta(\theta) = \phi\exp\{-\theta/\phi\}$, $d(y) = \exp\{-y/\phi\}$, $v(y) = -y/\phi$, $\hat{\theta} = -\phi\log\{n^{-1}\sum_{i=1}^{n}\exp(-y_i/\phi)\}$, $\alpha_0 = -1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = \alpha_0/2$, $\alpha_3 = -\alpha_1$.

(xviii) Truncated extreme value ($\theta > 0$, $y > 0$): $\alpha(\theta) = \theta^{-1}$, $\zeta(\theta) = \theta$, $d(y) = \exp\{y\} - 1$, $v(y) = y$, $\hat{\theta} = n^{-1}\{\sum_{i=1}^{n}\exp(y_i)\} - 1$, $\alpha_0 = 1/(3n^{1/2})$, $\alpha_1 = -1/(36n)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -7\alpha_1$.

(xix) Lognormal ($\theta > 0$, $-\infty < \mu < \infty$, $\mu$ known, $y > 0$): $\alpha(\theta) = \theta^{-2}$, $\zeta(\theta) = \theta$, $d(y) = (\log y - \mu)^2/2$, $v(y) = -\log y - \{\log(2\pi)\}/2$, $\hat{\theta} = [n^{-1}\sum_{i=1}^{n}\{\log(y_i) - \mu\}^2]^{1/2}$, $\alpha_0 = 4/\{6(2n)^{1/2}\}$, $\alpha_1 = -1/(18n)$, $\alpha_2 = -\alpha_0/4$, $\alpha_3 = -\alpha_0$.

(xx) Normal ($\theta > 0$, $-\infty < \mu < \infty$, $-\infty < y < \infty$):

(a) $\mu$ known: $\alpha(\theta) = (2\theta)^{-1}$, $\zeta(\theta) = \theta^{1/2}$, $d(y) = (y - \mu)^2$, $v(y) = -\{\log(2\pi)\}/2$, $\hat{\theta} = n^{-1}\sum_{i=1}^{n}(y_i - \mu)^2$, $\alpha_0 = 2^{1/2}/(3n^{1/2})$, $\alpha_1 = -1/(18n)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -7\alpha_1$.

(b) $\theta$ known: $\alpha(\mu) = -\mu/\theta$, $\zeta(\mu) = \exp\{\mu^2/(2\theta)\}$, $d(y) = y$, $v(y) = -\{y^2 + \log(2\pi\theta)\}/2$, $\hat{\mu} = \bar{y}$, $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0$.

(xxi) Inverse Gaussian ($\theta > 0$, $\mu > 0$, $y > 0$):

(a) $\mu$ known: $\alpha(\theta) = 1/\theta$, $\zeta(\theta) = \theta^{1/2}$, $d(y) = (y - \mu)^2/(2\mu^2 y)$, $v(y) = -\{\log(2\pi y^3)\}/2$, $\hat{\theta} = n[\sum_{i=1}^{n}\{(y_i - \mu)^2/(y_i\mu^2)\}]$, $\alpha_0 = 2^{1/2}/(3n^{1/2})$, $\alpha_1 = -1/(18n)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = -7\alpha_1$.

(b) $\theta$ known: $\alpha(\mu) = \theta/(2\mu^2)$, $\zeta(\mu) = \exp\{-\theta/\mu\}$, $d(y) = y$, $v(y) = -\theta/(2y) + [\log\{\theta/(2\pi y^3)\}]/2$, $\hat{\mu} = \bar{y}$, $\alpha_0 = \mu^{1/2}/\{2(n\theta)^{1/2}\}$, $\alpha_1 = \mu/(8n\theta)$, $\alpha_2 = -\alpha_0$, $\alpha_3 = 3\alpha_1$.

(xxii) McCullagh ($\theta > -1/2$, $-1 \le \mu \le 1$, $\mu$ known, $0 < y < 1$): $\alpha(\theta) = -\theta$, $\zeta(\theta) = 4^{-\theta}\,\mathrm{B}(\theta + 1/2, 1/2)$, $d(y) = \log[y(1 - y)/\{(1 + \mu)^2 - 4\mu y\}]$, $v(y) = -[\log\{y(1 - y)\}]/2$, $\hat{\theta}$ is obtained as solution of the equation $\psi(\hat{\theta} + 1/2) - \psi(\hat{\theta} + 1) = \log 4 - n^{-1}\sum_{i=1}^{n}\log[\{y_i(1 - y_1)\}/\{(1 + \mu)^2 - 4\mu y_i\}]$,

$$\alpha_0 = \frac{\{\psi''(\theta + 1) - \psi''(\theta + 0.5)\}}{6[n\{-\psi'(\theta + 1) + \psi'(\theta + 0.5)\}^3]^{1/2}},$$

$$\alpha_1 = -\frac{14\{\psi''(\theta + 1) - \psi''(\theta + 0.5)\}^2 + 9\{-\psi'(\theta + 0.5)\psi'''(\theta + 0.5)\}}{72n\{\psi'(\theta + 0.5) - \psi'(\theta + 1)\}^3} + \frac{9\{\psi'(\theta + 1)\psi'''(\theta + 0.5) + \psi'(\theta + 0.5)\psi'''(\theta + 1) - \psi'(\theta + 1)\psi'''(\theta + 1)\}}{72n\{\psi'(\theta + 0.5) - \psi'(\theta + 1)\}^3},$$

$$\alpha_2 = \frac{\psi''(\theta + 1) - \psi''(\theta + 0.5)}{3[n\{\psi'(\theta + 0.5) - \psi'(\theta + 1)\}^3]^{1/2}},$$

$$\alpha_3 = -\frac{4\{\psi''(\theta + 1) - \psi''(\theta + 0.5)\}^2 + 9\{-\psi'(\theta + 0.5)\psi'''(\theta + 0.5)}{72n\{\psi'(\theta + 0.5) - \psi'(\theta + 1)\}^3} + \frac{9\{\psi'(\theta + 1)\psi'''(\theta + 0.5) + \psi'(\theta + 0.5)\psi'''(\theta + 1) - \psi'(\theta + 1)\psi'''(\theta + 1)\}}{72n\{\psi'(\theta + 0.5) - \psi'(\theta + 1)\}^3},$$

where $B(\cdot,\cdot)$ is the beta function (see McCullagh, 1989).

(xxiii) von Mises ($\theta > 0$, $0 < \mu < 2\pi$, $\mu$ known, $0 < y < 2\pi$): $\alpha(\theta) = -\theta$, $\zeta(\theta) = 2\pi I_0(\theta)$, $d(y) = \cos(y-\mu)$, $v(y) = 0$, $\hat\theta = r^{-1}\{n^{-1}\sum_{i=1}^{n}\cos(y_i-\mu)\}$,

$$\alpha_0 = \frac{r''(\theta)}{6\{nr'(\theta)^3\}^{1/2}}, \quad \alpha_1 = \frac{-14r''(\theta)^2 + 9r'(\theta)r'''(\theta)}{72nr'(\theta)^3}, \quad \alpha_2 = 2\alpha_0,$$

$$\alpha_3 = \frac{-4r''(\theta)^2 + 9r'(\theta)r'''(\theta)}{72nr'(\theta)^3},$$

where $I_\nu(\cdot)$ is the modified Bessel function of first kind and $\nu$th order, and $r(\theta) = I_0'(\theta)/I_0(\theta)$.

(xxiv) Generalized hyperbolic secant ($-\pi/2 \le \theta \le \pi/2$, $r > 0$, $r$ known, $0 < y < 1$): $\alpha(\theta) = \theta$, $\zeta(\theta) = \pi(\sec\theta)^r$, $d(y) = -\log\{y/(1-y)\}/\pi$, $v(y) = -\log\{y(1-y)\}/2$, $\hat\theta = \arctan[(\pi r)^{-1}[1+\pi[\prod_{i=1}^{n}\{(y_i/(1-y_i))\}]^{1/n}]]$, $\alpha_0 = \sin(\theta)/\{3(nr)^{1/2}\}$, $\alpha_1 = \{9-10\sin(\theta)^2\}/(36nr)$, $\alpha_2 = \{2\sin(\theta)\}/\{3(nr)^{1/2}\}$, $\alpha_3 = \{9+10\sin(\theta)^2\}/(36nr)$.

It is possible to check by direct calculation that the formulae for the $\alpha$'s are correct for some distributions which have closed-form MLEs, namely: binomial, Poisson, gamma (case a), Pareto, Laplace, extreme value, normal (cases a and b) and inverse Gaussian (cases a and b). Clearly, the test of the mean of a normal distribution with known variance (case xx-b) is the only case for which $\alpha_i = 0$ for $i = 0,1,2,3$, confirming that in this case the MLE has an exact normal distribution. For the binomial case, $nm\hat\theta$ has a binomial distribution with parameters $nm\theta$ and the application of the classical Cornish–Fisher polynomial transformation (McCullagh, 1987, p.166) to the standardized statistic $S = (\hat\theta-\theta)[nm/\{\theta(1-\theta)\}]^{1/2}$ lead to the $\alpha$'s given in case (i). Analogously, we verify the $\alpha$'s for the Poisson and inverse Gaussian (case xxi-b) distributions. For the Poisson case, $n\hat\theta$ has a Poisson distribution with mean $n\theta$ and by applying the Cornish–Fisher transformation to $S = (\hat\theta-\theta)(n/\theta)^{1/2}$ we get the $\alpha$'s given above. For the inverse Gaussian distribution with known precision parameter $\theta$ (case xxi-b), the MLE $\hat\mu = \bar y$ of the mean $\mu$ has an inverse Gaussian distribution with parameters $\mu$ and $n\theta$, and then the $\alpha$'s follow from the Cornish–Fisher polynomial transformation of the standardized statistic $S = (\hat\mu-\mu)/\mu(n\theta/\mu)^{1/2}$.

For the gamma with known index $k$ and unknown mean $\theta$ (case x-a), Pareto, Laplace, extreme value, normal with known mean $\mu$ and unknown variance $\theta$ (case xx-a), and inverse Gaussian with known mean $\mu$ and unknown dispersion parameter $\theta$ (case xxi-a) distributions, the MLE $\hat\theta$ is proportional to a chi-squared random variable and we can easily obtain the $\alpha$'s from the Cornish–Fisher transformation applied to $S = (\chi_n^2 - n)/(2n)^{1/2}$. Thus, the adjusted variable $S^*$ defined by

$$S^* = S - \frac{\sqrt{2}}{3\sqrt{n}}(S^2-1) + \frac{7S^3 - S}{18n}$$

is asymptotically N(0,1) with an error $\mathcal{O}(n^{-3/2})$. A simple illustration is provided by taking the Laplace distribution with known location parameter $k$ (case xvi) for which $\hat\theta = \sum_i^n |y_i - k|/n \sim \theta/(2n)\chi_{2n}^2$. Defining $S = (\hat\theta-\theta)n^{1/2}/\theta \sim (\chi_{2n}^2 - 2n)/(4n)^{1/2}$, one easily obtains the adjusted statistic $S^*$.

# 5 Natural exponential family

In Section 3, we presented an improved standardized MLE $S^*$ for the pivotal quantity $S = (\hat{\theta} - \theta)\kappa_{\theta,\theta}^{1/2}$ in the one-parameter exponential family model (8). Our goal now is to derive an improved standardized estimate for the natural parameter $\alpha = \alpha(\theta)$. To that end, we assume that the model is parameterized in its natural form

$$\pi(y;\alpha) = \frac{1}{\delta(\alpha)} \exp\{-\alpha d(y) + v(y)\}, \tag{10}$$

where $-d(y)$ is the canonical statistic and $\log\{\delta(\alpha)\}$ is the cumulant generator of $-d(y)$. We now obtain a modified statistic $S^*$ for the canonical pivot $S$ whose distribution is standard normal when terms of order smaller than $n^{-1}$ are ignored. Here, $\hat{\alpha}$ is the MLE of $\alpha$, which comes from $n\beta'(\hat{\alpha}) = -\sum d(y_i)$, where $\beta = \beta(\alpha) = -\mathrm{E}\{d(y)\} = \mathrm{d}\log\{\delta(\alpha)\}/\mathrm{d}\alpha$ and $\kappa_{\alpha,\alpha} = n\beta'$ is the information for $\alpha$, where primes denote, from now on, derivatives with respect to $\alpha$. Let $\gamma_1^2 = \beta''^2/\beta'^3$ and $\gamma_2 = \beta'''/\beta'^2$ be the third and fourth standardized cumulants of $-d(y)$, respectively. These quantities are the usual measures of skewness and kurtosis of $-d(y)$. The coefficients $\alpha$'s which define the improved statistic $S^* = S + \sum_{i=0}^{3} \alpha_i S^i$ for $S$ were given at the end of Section 3.

When the cumulant generator function $\delta(\alpha)$ has the form

$$\delta(\alpha) = c_1 \exp(c_2\alpha)(k\alpha + c_3)^{-1/(6k)}, \tag{11}$$

where the $c$'s and $k$ are arbitrary constants, $1 - k/n$ is the Bartlett correction to the likelihood ratio statistic (see Cordeiro et al., 1995) and the $\alpha$'s reduce to

$$\alpha_0 = -(1/3)(6k/n)^{1/2}, \quad \alpha_1 = -k/(6n), \quad \alpha_2 = 2\alpha_0, \quad \alpha_3 = -19\alpha_1. \tag{12}$$

Note that for 12 of the 24 distributions considered in Section 4, the $\alpha$'s are constant and their cumulant generators satisfy (11). Thus, in all these cases, equations (12) hold and can be used to improve the standardized MLE $S$.

We now give simple expressions for the $\alpha$'s to improve the canonical pivot $S = (\hat{\alpha} - \alpha)\kappa_{\alpha,\alpha}^{1/2}$ under special families of variance functions. We begin with the power variance function defined by $\beta' = \beta^p/c_0$, where $p \leq 0$ or $p \geq 1$ and $c_0 > 0$. The domain of $\beta$ is $\mathbb{R}$ for $p = 0$ and $\mathbb{R}^+$ otherwise. Here, $p = 0, 1, 2, 3$ for the normal, Poisson, gamma, and inverse Gaussian distributions, respectively. Other values of $p$ define a number of distributions which have been classified by Jørgensen (1987). We obtain

$$\alpha_0 = \frac{p\beta^{(p-2)/2}}{6(nc_0)^{1/2}}, \quad \alpha_1 = \frac{p(4p-9)\beta^{p-2}}{72nc_0}, \quad \alpha_2 = 2\alpha_0, \quad \alpha_3 = \alpha_1. \tag{13}$$

Next, we consider the family of variance functions defined as polynomials of order less than or equal to 3, say $\beta' = c_0 + c_1\beta + c_2\beta^2 + c_3\beta^3$. When $c_3 = 0$, which corresponds to a quadratic variance function, Morris (1982) showed that there are only the following six distributions in the natural exponential family (10): normal, Poisson, binomial, negative binomial, gamma and generalized hyperbolic secant. All these distributions were studied in Section 4. Letac and Mora (1990) extended Morris's classification and showed that there are also only six distributions in (10) whose variance functions are polynomials of degree 3 in $\beta$ ($c_3 \neq 0$), namely: Abel, Takács, strict arcsine, large arcsine, Ressel and inverse Gaussian distributions (see

Ferrari et al., 1996, for further details). For the cubic variance function we obtain, using MAPLE,

$$\begin{aligned}
\alpha_0 &= \frac{c_1 + 2c_2\beta + 3c_3\beta^2}{6\sqrt{n(c_0 + c_1\beta + c_2\beta^2 + c_3\beta^3)}}, \\
\alpha_1 &= \frac{-5c_1^2 + 18c_0c_2 + 2\beta(27c_0c_3\beta - c_1c_2) + 2\beta^2(12c_1c_3 - c_2^2) + 12c_2c_3\beta^3 + 9c_3^2\beta^4}{72n(c_0 + c_1\beta + c_2\beta^2 + c_3\beta^3)}, \qquad (14) \\
\alpha_2 &= 2\alpha_0, \qquad \alpha_3 = \alpha_1.
\end{aligned}$$

Equations (14) apply to the 12 cases discussed above in order to derive improved canonical statistics.

We can also obtain, using MAPLE or MATHEMATICA, the $\alpha$'s for wider classes of variance functions, for example, the variance function $\beta' = P + Q\sqrt{R}$, where $P$, $Q$ and $R$ are polynomials in $\beta$ of degree less than or equal to 3, 2 and 2, respectively. This variance function includes the Babel class of variance functions for which $\beta' = a\Delta + (b\beta + c)\sqrt{\Delta}$, where $\Delta$ is a polynomial of degree less than or equal 2 in $\beta$ which is not a perfect square and $a$, $b$ and $c$ are three real numbers. However, the expressions of the $\alpha$'s for these variance functions are too cumbersome to be reported here.

## 6 Simulation results

In this section we present the results of two small Monte Carlo simulation experiments where we study the finite-sample distributions of $S$ and $S^*$. The first experiment uses an extreme value distribution with $\theta = \phi = 10$, where $\phi$ is assumed to be known and we are interested in the estimation of $\theta$. The second simulation involves a Weibull distribution with $\theta = \phi = 10$, where $\phi$ is known and we wish to estimate $\theta$. The total number of Monte Carlo replications was set at 20,000. All simulations were performed using Ox matrix programming language (Doornik, 1996), which uses an improved version of the Park and Miller (1988) uniform variate generator. Extreme value and Weibull variates were obtained as transformations of the generated uniform variates using the algorithms in Devroye (1986). In each simulation, the mean, variance, skewness and kurtosis of both the uncorrected ($S$) and the corrected ($S^*$) estimators were computed. The sample sizes considered range from 5 to 50 observations. Good agreement with asympotic theory happens when these figures are, on average, close to 0, 1, 0, and 3, respectively. The results for the extreme value and Weibull distributions are given in Tables 1 and 2. They show that the correction proposed in this paper is quite effective in bringing the distribution of the maximum likelihood estimator closer to its limiting normal distribution, even when the sample size is quite small. It is noteworthy that the correction tends to reduce the skewness of the distribution of the maximum likelihood estimator. That is, the finite-sample distribution of the maximum likelihood estimator tends to be skewed relative to its limiting distribution and our correction substantially reduces such skewness. The distribution of $S$ in the extreme value distribution is positevely skewed, whereas it tends to be negatively skewed in the Weibull case. In the first case (extreme value distribution), for example, the skewness of $S$ for 5 and 10 observations equals 0.47 and 0.31, whereas for $S^*$, the corrected estimator, one obtains 0.01 and $-0.01$, thus much closer to the asymptotic value of zero. In the second case (Weibull distribution), the corresponding skewness measures for the maximum likelihood estimator are $-0.32$ and $-0.23$, whereas for the corrected estimator one obtains 0.00 and $-0.01$, respectively. The mean of the corrected estimator is also closer to the asymptotic value of zero. When $n = 5, 10$, the

Table 1. Simulation Results I

| $n$ | estimator | mean | variance | skewness | kurtosis |
|---|---|---|---|---|---|
| 5 | $S$ | 0.23 | 1.11 | 0.47 | 3.44 |
| | $S^*$ | 0.00 | 1.00 | 0.01 | 3.06 |
| 10 | $S$ | 0.16 | 1.06 | 0.31 | 3.17 |
| | $S^*$ | −0.01 | 1.00 | −0.01 | 3.00 |
| 15 | $S$ | 0.13 | 1.04 | 0.25 | 3.08 |
| | $S^*$ | −0.01 | 1.01 | −0.01 | 2.98 |
| 20 | $S$ | 0.11 | 1.04 | 0.24 | 3.12 |
| | $S^*$ | 0.00 | 1.02 | 0.01 | 3.01 |
| 25 | $S$ | 0.10 | 1.03 | 0.21 | 3.13 |
| | $S^*$ | −0.01 | 1.01 | 0.00 | 3.05 |
| 30 | $S$ | 0.09 | 1.03 | 0.21 | 3.22 |
| | $S^*$ | 0.00 | 1.01 | 0.02 | 3.11 |
| 35 | $S$ | 0.09 | 1.02 | 0.20 | 3.13 |
| | $S^*$ | 0.00 | 1.01 | 0.02 | 3.05 |
| 40 | $S$ | 0.08 | 1.03 | 0.18 | 3.08 |
| | $S^*$ | 0.00 | 1.01 | 0.02 | 3.01 |
| 45 | $S$ | 0.08 | 1.03 | 0.15 | 3.03 |
| | $S^*$ | 0.00 | 1.01 | 0.00 | 2.99 |
| 50 | $S$ | 0.07 | 1.02 | 0.11 | 3.01 |
| | $S^*$ | 0.00 | 1.02 | −0.03 | 3.00 |

estimated means of the uncorrected estimator under the extreme value (Weibull) distribution are 0.23 and 0.16 (−0.20 and −0.14), whereas the estimated means of the corrected estimator are 0.00 and −0.01 (0.01 and 0.01), respectively. We then note that all four cumulants of the corrected estimator are generally closer to their limiting values. In short, the correction appears to be effective, even when the sample size is quite small.

## Acknowledgements

The first three authors acknowledge the financial support from CNPq/Brazil, and the fourth author acknowledges the financial support from the Office of Research Development Administration at Southern Illinois University.

## References

Abell, M.L. and Baselton, J.P. (1994). *The MAPLE V Handbook.* New York: AP Professional.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

Barndorff-Nielsen, O.E. (1988). Discussion of "Saddlepoint methods and statistical inference" by N. Reid.

Table 2. Simulation Results II

| $n$ | estimator | mean | variance | skewness | kurtosis |
|---|---|---|---|---|---|
| 5 | $S$ | −0.20 | 1.07 | −0.32 | 3.08 |
| | $S^*$ | 0.01 | 1.01 | 0.00 | 3.04 |
| 10 | $S$ | −0.14 | 1.03 | −0.23 | 3.08 |
| | $S^*$ | 0.01 | 1.00 | 0.01 | 2.98 |
| 15 | $S$ | −0.11 | 1.02 | −0.16 | 3.07 |
| | $S^*$ | 0.00 | 1.00 | 0.02 | 3.02 |
| 20 | $S$ | −0.10 | 1.01 | −0.12 | 3.00 |
| | $S^*$ | 0.00 | 1.00 | 0.04 | 2.98 |
| 25 | $S$ | −0.09 | 1.01 | −0.11 | 2.99 |
| | $S^*$ | 0.00 | 1.00 | 0.03 | 2.96 |
| 30 | $S$ | −0.08 | 1.01 | −0.09 | 2.99 |
| | $S^*$ | 0.00 | 1.00 | 0.04 | 2.99 |
| 35 | $S$ | −0.08 | 1.01 | −0.09 | 3.00 |
| | $S^*$ | 0.00 | 1.00 | 0.03 | 2.95 |
| 40 | $S$ | −0.08 | 1.01 | −0.08 | 2.95 |
| | $S^*$ | 0.00 | 1.01 | 0.02 | 2.94 |
| 45 | $S$ | −0.07 | 1.01 | −0.10 | 2.97 |
| | $S^*$ | −0.01 | 1.00 | 0.00 | 2.95 |
| 50 | $S$ | −0.07 | 1.02 | −0.09 | 3.00 |
| | $S^*$ | 0.00 | 1.01 | −0.01 | 2.99 |

*Statistical Science* **3**, 228-229.

Barndorff-Nielsen, O.E. (1990). Approximate interval probabilities. *Journal of the Royal Statistical Society* B **52**, 485-496.

Bhattacharya, R.N. and Rao, R.R. (1976). *Normal Approximation and Asymptotic Expansion.* New York: Wiley.

Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics* **6**, 434-451.

Burr, I.W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics* **13**, 215-232.

Cordeiro, G.M., Cribari-Neto, F., Aubin, E.C.Q., and Ferrari, S.L.P. (1995). Bartlett corrections for one-parameter exponential family models. *Journal of Statistical Computation and Simulation* **53**, 211-231.

Cordeiro, G.M. and Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order $n^{-1}$. *Biometrika* **78**, 573-582.

Cordeiro, G.M. and Ferrari, S.L.P. (1997). A general method for approximating to the distribution of some statistics. Submitted to *Communications in Statistics: Theory and Methods.*

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Cribari–Neto, F. and Cordeiro, G.M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339-367.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer-Verlag.

Doornik, J.A. (1996). *Ox: An Object–Oriented Matrix Programming Language.* London: International Thomson Business Press and `http://www.nuff.ox.ac.uk/Users/Doornik/`.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol.2, 2nd ed. New York: Wiley.

Ferrari, S.L.P., Botter, D.A., Cordeiro, G.M. and Cribari-Neto, F. (1996). Second- and third-order bias reduction for one-parameter family models. *Statistics and Probability Letters* **30**, 339-345.

Ferrari, S.L.P., Cordeiro, G.M., Uribe-Opazo, M.A. and Cribari-Neto, F. (1996). Improved score tests for one-parameter exponential family models. *Statistics and Probability Letters* **30**, 61-71.

Hill, G.W. and Davis, A.W. (1968). Generalized asymptotic expansions of Cornish-Fisher type. *Annals of Mathematical Statistics* **39**, 1264-1273.

Jensen, J.L. (1992). The modified signed log likelihood statistic and saddlepoint approximation. *Biometrika* **79**, 693-704.

Johnson, N.L., Kotz, S. and Balakrishnan (1994). *Continuous Univariate Distributions*, vol.1, 2nd ed. New York: Wiley.

Johnson, N.L., Kotz, S. and Balakrishnan (1995). *Continuous Univariate Distributions*, vol.2, 2nd ed. New York: Wiley.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions*, 2nd ed. New York: Wiley.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society* B **49**, 127-162.

Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **71**, 233-244.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance function. *Annals of Statistics* **18**, 1-37.

McCullagh, P. (1987). *Tensor Methods in Statistics.* London: Chapman and Hall.

McCullagh, P. (1989). Some statistical properties of a family of continuous univariate distributions. *Journal of the American Statistical Association* **84**, 125-141.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Annals of Statistics*

**9711 - YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H.** Stochastic Monotonicity properties of bayes estimation of the population size for capture-recapture data. IME-USP. 1997. 12p. (RT-MAE-9711)

**9712 - GUIMENEZ,P.;COLOSIMO,E.A.; BOLFARINE,H.** Asymptotic relative efficiency of wald tests in measurement error models. IME-USP. 1997. 14p. (RT-MAE-9712)

**9713 - SCHONMANN, R.H.; TANAKA, N.I.** Lack of monotonicity in ferromagnetic ising model phase diagrams. IME-UPS. 1997. 12p. (RT-MAE-9713)

**9714 - LEITE,J.G.; TORRES, V.H.S.; TIWARI, R.C.; ZALKIKAR, J.** Bayes estimation of dirichlet process parameter. IME-USP. 1997. 21p. (RT-MAE-9714)

**9715 - GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.** Hypotheses testing based on a corrected score function for errors-in-variables models. 1997. 16p. (RT-MAE-9715)

**9716 - PAULA, G.A.; BOLFARINE,H.** Some results on the slope of the linear regression model for the analysis of pretest-posttest data. 1997. 08p. (RT-MAE-9716)

**9717 - AUBIN, E.C.Q.; CORDEIRO, G.M.** Some adjusted likelihood ratio tests for heteroscedastic regression models. 1997. 19p. (RT-MAE-9717)

**9718 - ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a one-stage dose-response model using the minimum sum absolute errors estimators. 1997. 09p. (RT-MAE-9718)

**9719 - ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a two-stage dose-response model using the minimum sum of absolute errors estimators. 1997. 15p. (RT-MAE-9719)

**9720 - COLLET, P.; GALVES, A.; SCHMITT, B.** Fluctuations of repetition time for gibbsian sources. 1997. 08p. (RT-MAE-9720)

**9721 - CORDEIRO, G.M.; FERRARI, S.L.P.;CYSNEIROS, A.H.M.A** A formula to improve score test statistics. 1997. 10p. (RT-MAE-9721)

**9722 - ZUAZOLA, P.I.;GASCO-CAMPOS,L.; BOLFARINE, H.** Pearson type II measurement error models. 1997. 24p. (RT-MAE-9722)

**9723 - BUENO, V.C.** A note on asymptotics reliability of a linearly connected system. 1997. 07p. (RT-MAE-9723)

**9724 - NARULA, S.C.; SALDIVA, P.H.N; ANDRÉ, C.D.S.; ELIAN, S.N.; FERREIRA, A.F.; CAPELOZZI, V.** The minimum sum of absolute errors regression: an overview. 1997. 22p. (RT-MAE-9724)

**9725 - BOLFARINE, H.; LIMA, C.R.O.P.; SANDOVAL, M.C.** Linear calibration in functional regression models with one of the variances known. 1997. 27p. (RT-MAE-9725)

**The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.**


*Departamento de Estatística*
*IME-USP*
*Caixa Postal 66.281*
*05315-970 - São Paulo, Brasil*